

Appln No. 10/524,606
Amdt date December 10, 2007
Reply to Office action of September 10, 2007

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1-75. (Cancelled)

76. (New) In a training data set D containing gene expression data for a plurality of genes derived from a normal cell group including a plurality of normal cells and a diseased cell group including a plurality of diseased cells associated with a disease, a method for determining a plurality of emerging patterns within the gene expression data, wherein said emerging patterns are configured to be used to test cells from a test sample for the presence of the disease, the method comprising the steps of:

determining a plurality of emerging patterns from the training data set D, wherein each of said emerging patterns comprises at least one condition based on a fixed range of said gene expression data for at least one of said plurality of genes, and wherein a plurality of occurrences in one of said normal cell group and diseased cell group satisfies said at least one condition, but no occurrence in the other one of said normal cell group and diseased cell group satisfies said at least one condition; and

creating two lists, wherein a first list of said two lists contains a frequency of occurrence, $f_1(m)$, of each emerging pattern $EP_1(m)$ from said plurality of emerging patterns that has a non-zero occurrence in data associated with said normal cells, and a second list of said two lists contains a frequency of occurrence, $f_2(m)$, of each emerging pattern $EP_2(m)$ from said plurality of emerging patterns that has a non-zero occurrence in data associated with said diseased cells,

wherein said two lists each provide a plurality of emerging patterns configured to be used as a diagnostic tool to determine whether the cells from the test sample are normal or diseased.

Appln No. 10/524,606
Amdt date December 10, 2007
Reply to Office action of September 10, 2007

77. (New) The method of claim 76, wherein at least one of said emerging patterns from the training data set D comprises at least three conditions for each of said normal and diseased cells.

78. (New) The method of claim 76, further comprising steps for:
providing a test sample;
testing a plurality of genes in said test sample to produce a test data set, said plurality of genes in said test sample comprising at least a portion of the genes identified in said plurality of emerging patterns for said training data set D;
determining a plurality of emerging patterns from the test data set, wherein each of said emerging patterns from the test data set comprises at least one condition based on a fixed range of said gene expression data for at least one of said plurality of test genes, and wherein a plurality of occurrences in at least one of said normal cell group and diseased cell group satisfies said at least one condition, but no occurrence in the other one of the normal cell group and diseased cell group satisfies said at least one condition;
identifying a fixed number, k , of emerging patterns extracted from said training data, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns for said training data set D;
calculating two scores, wherein a first score corresponding to the data associated with the normal cells is derived from the frequencies of k emerging patterns in said first list that also occur in said test data, and a second score corresponding to the data associated with the diseased cells is derived from the frequencies of k emerging patterns in said second list that also occur in said test data; and
determining whether said test cells are normal or diseased based on said two scores.

79. (New) The method of claim 78, additionally comprising, prior to said determining step, a step for applying an entropy based discretization technique to said training data set to generate a cut point that defines said fixed range, such that said normal data falls

within said fixed range on one side of said cut point, and said diseased data falls on an opposite side of said cut point.

80. (New) The method of claim 79, additionally comprising applying a method of correlation based feature selection to said training data set, after said discretizing.

81. (New) The method of claim 79, additionally comprising applying a chi-squared method to said training data set, after said discretizing.

82. (New) The method of claim 79, wherein said emerging patterns are jumping emerging patterns.

83. (New) A system for determining a plurality of emerging patterns within a training data set D containing gene expression data for a plurality of genes derived from a normal cell group including a plurality of normal cells and a diseased cell group including a plurality of diseased cells associated with a disease, wherein said emerging patterns are configured to be used to test cells from a test sample for the presence of the disease, the system comprising:

at least one memory, at least one processor, and at least one user interface, all of which are connected to one another by at least one bus;

wherein said at least one processor is configured to:

access the gene expression data within the training data set D;

determine a plurality of emerging patterns from the training data set D, wherein each of said emerging patterns comprises at least one condition based on a fixed range of said gene expression data for at least one of said plurality of genes, and wherein a plurality of occurrences satisfies said at least one condition for one of said normal cell group and diseased cell group, but no occurrence satisfies said at least one condition for the other one of said normal cell group and diseased cell group; and

create two lists, wherein a first list of said two lists contains a frequency of

occurrence, $f_1(m)$, of each emerging pattern $EP_1(m)$ from said plurality of emerging patterns that has a non-zero occurrence in data associated with said normal cells, and a second list of said two lists contains a frequency of occurrence, $f_2(m)$, of each emerging pattern $EP_2(m)$ from said plurality of emerging patterns that has a non-zero occurrence in data associated with said diseased cells,

wherein said two lists each provide a plurality of emerging patterns configured to be used as a diagnostic tool to determine whether the cells from the test sample are normal or diseased.

84. (New) The system of claim 83, wherein at least one of said emerging patterns from the training data set D comprises at least three conditions for each of said normal and diseased cells.

85. (New) The system of claim 83, wherein said processor is further configured to:
receive data from a plurality of genes in said test sample to produce a test data set, said plurality of genes in said test sample comprising at least a portion of the genes identified in said plurality of emerging patterns for said training data set D;

determine a plurality of emerging patterns from the test data set, wherein each of said emerging patterns from the test data set comprises at least one condition based on a fixed range of said gene expression data for at least one of said plurality of test genes, and wherein a plurality of occurrences in at least one of said normal cell group and diseased cell group satisfies said at least one condition, but no occurrence in the other one of the normal cell group and diseased cell group satisfies said at least one condition;

identify a fixed number, k , of emerging patterns extracted from said training data, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns for said training data set D;

calculate two scores, wherein a first score corresponding to the data associated with the normal cells is derived from the frequencies of k emerging patterns in said first list that also

Appln No. 10/524,606
Amdt date December 10, 2007
Reply to Office action of September 10, 2007

occur in said test data, and a second score corresponding to the data associated with the diseased cells is derived from the frequencies of k emerging patterns in said second list that also occur in said test data; and

determine whether said test cells are normal or diseased based on said two scores.

86. (New) The system of claim 85, wherein, prior to determining said plurality of emerging patterns, said processor is configured to apply an entropy based discretization technique to said training data set to generate a cut point that defines said fixed range, such that said normal data falls within said fixed range on one side of said cut point, and said diseased data falls on an opposite side of said cut point.

87. (New) The system of claim 86, wherein, after applying said discretization method, said processor is further configured to apply a method of correlation based feature selection to said training data set.

88. (New) The system of claim 86, wherein, after applying said discretization method, said processor is further configured to apply a chi-squared method to said training data set.

89. (New) The system of claim 86, wherein said emerging patterns are jumping emerging patterns.